Quantitative approaches to content analysis

Identifying conceptual drift across publication outlets

Indulska, Marta; Hovorka, Dirk S.; Recker, Jan

Published in:
European Journal of Information Systems

DOI:
10.1057/ejis.2011.37

Quantitative Approaches to Content Analysis: Identifying Conceptual Drift across Publication Outlets

Abstract

Unstructured text data, such as emails, blogs, contracts, academic publications, organizational documents, transcribed interviews, and even tweets, are important sources of data in Information Systems research. Various forms of qualitative analysis of the content of this data exist and have revealed important insights. Yet, to date, these analyses have been hampered by limitations of human coding of large data sets, and by bias due to human interpretation. In this paper, we compare and combine two quantitative analysis techniques to demonstrate the capabilities of computational analysis for content analysis of unstructured text. Specifically, we seek to demonstrate how two quantitative analytic methods, *viz.*, Latent Semantic Analysis and data mining, can aid researchers in revealing core content topic areas in large (or small) data sets, and in visualizing how these concepts evolve, migrate, converge or diverge over time. We exemplify the complementary application of these techniques through an examination of a 25-year sample of abstracts from selected journals in Information Systems, Management and Accounting disciplines. Through this work, we explore the capabilities of two computational techniques, and show how these techniques can be used to gather comprehensive insights from a large corpus of unstructured text.

Keywords

Unstructured data analysis, quantitative semantic analysis, text mining

Introduction

Increasingly, unstructured information from academic and trade publications, organizational reports, marketing materials, websites, blogs, email, meeting notes, contracts, organizational policies, and conversation transcripts are created, stored, or transmitted via information systems. This largely unstructured text data represents a major research opportunity for the Information Systems (IS) discipline, as it has the potential to provide insight into phenomena that involve verbal and/or written communications. Until recently, however, content analysis or classification of large volumes of data has been a time consuming and resource intensive task. One of the most prevalent issues with content analysis is the reliance on human coders

as well as the necessity of pre-defined dictionaries of concepts or terms (e.g., Smith & Humphreys, 2006). This situation brings forward at least three types of limitations. First, human coding of any sort of unstructured text is susceptible to subjectivity in the analysis and requires investments in inter-coder reliability testing. Second, the human analysis of text data is prone to variability of human categorization of research topics/keywords (Kruschke, 1992; Hampton, 1995). Pre-defined dictionaries induce bias into the analysis, restrict the exploration of material to a limited scope, and limit the possibility of having new concepts emerge from the material. Third, interpretation of the data is prone to subjective interpretation bias, reducing the external validity of such research. While methodological guidelines have been offered to assist with these interpretative tasks (Urquhart et al., 2010), e.g., by offering different types of codes (Miles & Huberman, 1994) or the use of coding teams (Fernandez, 2004), still, the subjective bias introduced by the researcher(s) cannot fully be alleviated.

To address such limitations, *quantitative computational approaches* towards text analysis from information and cognitive science have gained prominence as valid methods for facilitating the examination of text corpi (Landauer, 2007; Larsen et al., 2008; Sidorova et al., 2008). Once data is cleaned and prepared, computational analysis takes significantly less time than manual data coding and reduces the bias in human interpretation and categorization of text data, which can vary across multiple coders, or over an extended time period. Given the reduction in resources required to analyze large data sets, researchers are enabled to perform analyses with different parameters, or to perform multiple analyses of data alongside different dimensions (e.g., longitudinal changes in meaning over time *vs.* aggregate meaning, splitting of texts into different unit sizes, or seeding the analysis to look for specific topics in the text). Such additional analyses and use of different parameters has the potential to result in richer and more relevant research outcomes.

We argue that computational approaches can complement, rather than displace, the human interpretation and analysis of large corpi of text. This is because computational approaches can analyse textual data sets in a repeatable manner by identifying and relating important terms, concepts and relationships. These outputs may then be subjected to additional statistical or visual descriptions and the outputs may in turn facilitate additional detailed analysis through human interpretation (Weber, 1990). In essence, the strength of such approaches lies in the provision of a analytic process that reproducible and is capable of handling data volumes that may be problematic for human analysis. Computational

approaches have advantages that relate to scalability, repeatability, and consistency. Although these techniques are sufficient for many analytic purposes, they may also be used in a complementary manner with additional human analysis, for instance, to identify changes of context, emotion, or tone that are currently problematic for computational analysis.

Quantitative computational approaches have been applied to phenomena in Management, Sociology, Marketing, Health, and other disciplines – for example in literature indexing (Foltz, 1995), evaluation of medical interventions and diagnoses (Elvevåg et al., 2007; Al Qenaei, 2009), and language-based communication (Dong, 2005). Still, these applications were restricted to one analysis approach in isolation, leading to a gap in knowledge about the similarities and differences in the examination of data between different analytical techniques. Moreover, with a few exceptions (Davies et al., 2006; Larsen et al., 2008; Sidorova et al., 2008), the use of quantitative computational approaches to content analysis in Information Systems research remains a largely underexplored area (King, 2009), which is surprising given the vast quantities of data relating to all aspects of modern life. The value of such approaches for the analysis of large sets of data pertaining to the development, use or impact of Information Systems is significant. For example, analysis of communications in IT service call centers can identify structural problems in products or services, while analyses of user complaints can lead to prioritization of systems maintenance requests. Examinations of technicians' service reports may identify emergent knowledge that can lead to the identification of best practices by filtering out irrelevant passages in the reports. Other application areas include knowledge management (Bobrow & Whalen, 2002), expert systems (Aniba et al., 2009), customer relationship management (Coussement & van den Poel, 2008) or online communication (Penn-Edwards, 2010). The quantitative output of computational approaches to content analysis can be used in factor analysis, clustering and other statistical techniques to determine relationships among groups of text documents. Areas include document classification, text summarization, information visualization and other applications requiring extraction of meaning from text units.

In this research we illustrate the complementary use of two different computational semantic analysis techniques to examine a large textual data set. The main aim of the paper is to introduce, explore and exemplify Latent Semantic Analysis (LSA) and data mining as complementary approaches for quantitative analysis of text data. To illustrate the application of these approaches, we use them to analyze an exploratory sample of journals identified by Trieschmann *et al*. (2000) as core journals for different business disciplines. Specifically, we

examine 8544 abstracts from journals in Information Systems, Management, and Accounting over a period of 25 years. We show how LSA can be used to identify the relative *conceptual drift* among the aggregate set of journals from each of the three academic disciplines, and how data mining can be used to identify the core concepts that are involved in, and relevant to, the noted conceptual drift among the selected journal sets. The complementary application of the two computational analysis approaches represents a new application of quantitative analysis of textual data in that it does not report a static analysis of meaning for a text corpus but rather the longitudinal changes in the corpus.

Background

Content analysis (e.g., Weber, 1990) is concerned with the semantic analysis of a body of text, to uncover the presence of strong concepts. In general, content analysis approaches fall into two major categories (Smith & Humphreys, 2006), *viz.* conceptual and relational. In *conceptual analysis*, text material is examined for the presence, frequency and centrality of concepts. Such concepts can represent words, phrases, or more complex definitions. *Relational analysis*, on the other hand, tabulates not only the frequency of concepts in the body of text, but also the co-occurrence of concepts, thereby examining how concepts (pre-defined or emergent) are related to each other within the documents, for instance, in terms of affect extraction[1], contextual proximity or cognitive mapping. In this paper we are interested in *both* conceptual and relational analysis, specifically in the identification of central concepts within and across text corpi, and in the co-occurrence relationships between the concepts.

Content analysis was performed typically through involvement of trained human analysts who tagged corpi of text with pre-defined or emerging codes, thus introducing a source of bias before the coded data can even be properly analyzed. More recently, however, computational approaches have become available that allow experts, as well as novices, to explore the content of large bodies of text. Several such approaches have been developed, including hyperspace analogue to language (Burgess & Lund, 1997), latent semantic analysis (Landauer et al., 1998) and data mining tools such as Leximancer (Smith & Humphreys, 2006). Hyperspace analogue to language approaches are restricted in that they assume

[1] Affect extraction concerns the examination of text-based conversational material to uncover information about affect conveyed in the conversations. These affects could include emotions or moods (such as embarrassment, hostility), or evaluations (of goodness, importance, etc.).

symmetric relationships and close term co-occurrence (Stockwell et al., 2009), which is why we focus on Latent Semantic Analysis and data mining based approaches.

Latent Semantic Analysis

LSA is a content analysis method that uncovers latent semantic relationships within a corpus of text through statistical computations, to extract a quantification of the meaning of text units. The basis for this analysis is that the totality of information about all the word contexts, in which a given word does or does not appear, provides a set of mutual constraints that largely determines the similarity of meaning of words and of a set of words to each other (Landauer, 2007).

LSA calculates the relative positions of units of text in an n-dimensional semantic space. The size and boundary of a text unit is determined by the researcher and may range from the document abstract, to the entire document or individual paragraphs or sets of sentences. Interpretation of results from single-sentence text units, however, is questionable. LSA begins by treating each text unit as a 'bag of words' without structure. Non-content bearing words that occur only once, or almost always, in a text unit, (e.g. "of", "the", "and", "be") are removed. Articles, prepositions, pronouns, and conjunctions, as well as common adjectives, adverbs, and proper names may also be removed. The remaining words are stemmed or lemmatized to avoid morphological variants represented in the analysis. For example removing an "s" or "es" will convert some plurals to singulars, and stemming the words "uses," "using," and "used" reduces all three to "use", thus increasing identification of conceptually similar words and reducing processing time. Approaches, such as the Porter stemming algorithm (Porter, 1980), used here are automatic but need to be done carefully to avoid conflating similar terms and thus biasing the results towards greater convergence. Additionally, LSA document-term matrices are used to analyse terms within the context of the text artefact. The term and its co-occurring terms are used to calculate position in the semantic space such that "the aggregate of all the word contexts in which a given word does and does not appear provides a set of constraints that determines the similarity of meanings of words, and sets of words, to each other" (Landauer et al., 1998, p. 260).

The technical details of the subsequent numeric transformation and statistical processing steps are available in Larsen and Monarchi (2004). In brief, the method creates a sparse matrix with the unique stems as rows, the text units as columns, and the number of occurrences of a specific stem in a specific text as the cell value. In this research, the matrix

was weighted using a TFIDF weighting scheme (Term Frequency-Inverse Document Frequency – a statistical procedure that determines how relevant a particular word is to a particular artifact), before it was subjected to singular value decomposition (SVD). Alternative weighting schemes may be used, but our goal was not to compare results among the alternatives.[2] SVD creates a high-dimensional space in which each text unit occupies a specific location identified by its vector. The vector computed for each text unit is a cardinal number, which can be subjected to clustering algorithms, factor analysis, or other statistical techniques.

By aggregating the texts, a centroid for those text units can be located in the n-dimensional space representing that collection of texts (for example, all abstracts published in the selected journals from each specific discipline over 25 years). This representation allows a measurement of distance (by cosine of the angle or by Euclidian distance) between the collections of text (e.g., the centroid of each set of aggregated abstracts from each discipline in our case) to the centroid representing other aggregated abstracts. For illustration purposes, Figure 1 shows a two-dimensional representation of such an n-dimensional semantic space.

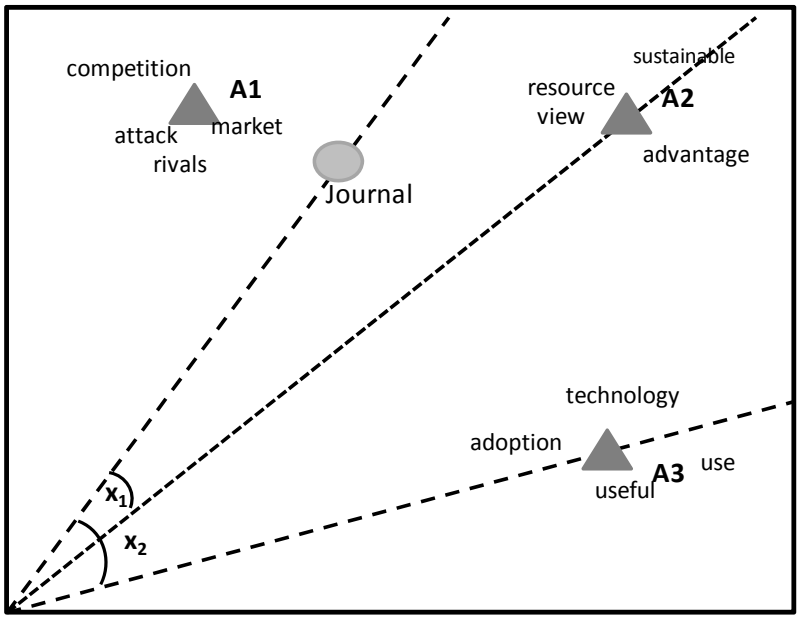


Figure 1: Two-dimensional representation of an n-dimensional semantic space using LSA

[2] We refer to Kontostathis and Pottenger (2006) for such a discussion. While we discuss in this paper a standard application of LSA, we note that different weighting schemes can also be used and that the results can also undergo further clustering. We refer the reader to Kontostathis and Pottenger (2006) for examples of different LSA settings.

In Figure 1, the symbol labeled 'Journal' represents the position of the semantic meaning from an aggregated set of abstracts published in the journal over a time period. The symbols labeled A1, A2, and A3 are article abstracts that represent the relative position of three abstracts in the same semantic space. The terms (words) around each abstract symbol are the four words that contribute most to the meaning of the abstract. From this hypothetical example, we can see that A1 and A2 are more similar in meaning (the distance as measured by the angle between them is smaller) to the journal (x_1) than A3 is to the journal (x_2). This suggests that A1 and A2 are more typical of topics areas published in the journal during the time period sampled.

Data Mining

Data mining is an automated approach to identifying patterns in sets of data. While many different data mining algorithms exist, applications suitable for text content analysis typically follow a three-step approach of (a) parsing text; (b) identification of concepts; and (c) clustering. Instead of implementing our own data mining algorithms, in our work we use an off-the-shelf data mining tool that is specialized for content analysis. While a number of such tools are available (e.g. IBM Content Analyzer, Lucene, SAS Text Miner, etc), we choose Leximancer[3], which we describe in the following pages. Our reasons for the selection are based on a number of arguments. First, we select Leximancer because prior research has demonstrated reproducibility of outcomes, as well as correlative and functional validity of the underlying algorithms (Smith & Humphreys, 2006). Selecting other tools/algorithms or developing one from scratch would require us to invest time in examining whether the output is comparable to human coding. Second, Leximancer features an iterative learning approach to coding concepts and themes in the data without the bias introduced by static dictionaries or thesauri. Third, it provides useful visualization of coded themes and concepts together with an interface that allows the researcher to drill down to the underlying data to understand the context of the concepts/themes, identify the sources and common discussion of the concepts, and identify relevant data files. Finally, Leximancer has been used in several academic studies on content analysis and appears to be the favored data mining-based content analysis tool (e.g., Davies et al., 2006; Martin & Rice, 2007; Stockwell et al., 2009).

[3]http://www.leximancer.com

Leximancer is used to analyze the content of collections of textual documents and to explore the extracted information statistically and visually. It performs full content analysis in two steps. First, during semantic extraction, Leximancer discovers concepts from the text data, with no requirement for a pre-defined dictionary (although one can be used if desired) or limitation of a specific spoken language. Leximancer defines a concept as a term that is less frequently yet consistently used with a more common term (Stockwell et al., 2009) – i.e. a concept is a collection of words that occurs together often. Common stop-words (such as "the", "an" etc) are ignored based on a default stop-word list for the English language (see Appendix A for a subset list of stop words as an example). Leximancer identifies these concepts using a Bayesian co-occurrence metric (Salton, 1989) to measure co-occurrence relevance. The computation uses a concept bootstrapping algorithm developed from a word sense disambiguation algorithm to identify families of weighted terms that tend to appear together in text (Yarowsky, 1995).

In a second step, called relational extraction, the emerging concepts are coded into the text and a thesaurus of terms is associated with each concept to classify text segments. The concepts are formed from correlated 'evidence words'. Consider, for example, that the concept of "system" comprises the evidence words "information", "systems", "system", "computer-based", "MIS", to name just a few. The evidence words for a concept are discovered though an iterative algorithm (typically over 2000-3000 iterations to ensure stability of analysis). Once the optimal weighted set of evidence words is found for each concept (i.e., the concept is stable), it is used to identify the concepts present in fragments of related text. In other words, each concept (which is an aggregation of its evidence words) has other concepts that it attracts (or is highly associated with contextually) as well as concepts that it repels (or is highly disassociated with contextually). The relationships are measured by the weighted sum of the number of times two concepts (i.e. the evidence words underlying these concepts) are found in the same block of text. The relation extraction algorithm is used to determine the confidence and relevancy of the terms to others in a specific block and across blocks. From this information, a matrix of concept co-occurrence is computed. On the diagonal of this matrix is the occurrence count for the relevant concepts. Each column of the co-occurrence matrix is divided by its diagonal, which results in each cell representing the probability of occurrence of the row term given the occurrence of the column term. The matrix is then used as a basis for the development of the core Leximnacer output – the concept map. The concept map depicts the forces (attract and repel) between the concepts

through distance between concepts on the map. Each of the identified concepts is placed on the map in proximity (i.e., the co-occurrence) to other concepts in the map through a derived combination of the direct and indirect relationships between those concepts. The concept map facilitates easy exploration of the data through drill-downs and hyperlinks such that the source of the concepts can be explored. Concepts are grouped into themes, which are clusters of frequently co-occurring concepts.

Further details about the algorithms underlying Leximancer are available in Smith and Humphreys (2006) and Stockwell et al. (2009).

Interpreting Leximancer Output

Leximancer generates several outputs that enable the researcher to judge the relevance of concept clusters (themes), the frequency of concepts relative to other concepts, concept connectedness ordered lists, concept co-occurence matrices (on which the maps are based), and others. Some of these outputs, e.g., the co-occurrence matrix, can be extracted and subjected to external analyses. However, the main interactive output of the Leximancer data mining tool is the aforementioned concept map, which is based on the co-occurrence matrix and is a visual representation of the core concepts and their interrelationships. Accordingly, in the following pages we provide an introduction to interpreting concept maps. For more details of other Leximancer output we refer the reader to explanations provided in Stockwell et al. (2009).

We introduce the reader to Leximancer concept maps through the example shown in Figure 2, which provides different representations of the same concept map generated from a 1977-1981 sample of MIS Quarterly journal abstracts. Perusing this example, we can highlight some of the features above and explain the basic concept map interpretation rules as follows:

- A concept map visualizes a collection of concepts. Concepts are represented by labeled and color coded dots. The labels (concept names) are the abbreviated single word descriptors for a collection of evidence words that make up the concept (e.g. refer to the “system” concept example above).
- The size and the brightness of a concept dot on the map is indicative of the concept’s strength within the body of analyzed text (i.e. the brighter, bigger the concept, the more often it – through its evidence words - appears in the text);

- The thickness and brightness of connections between concepts is indicative of the frequency of co-occurrence of the two concepts, as per the co-occurrence matrix. Two concepts that occur together frequently (through their underlying evidence words) will be connected by a thicker and brighter link than two concepts that co-occur less frequently.
- The relative distance of concepts on the map is indicative of the degree to which the concepts (through their underlying evidence words) appear together in the text (i.e. the concepts co-occur more frequently with concepts that are placed closer on the map, and less frequently with concepts that are placed further away).
- Concepts on a concept map are clustered based on their co-occurrence, thus representing themes within the analyzed text. Themes are formed around concepts that are highly connected and are automatically named after the strongest concept in the cluster (or theme).. The color used to represent the theme is indicative of the degree of connectedness of the concepts within the theme, with colors towards the red color spectrum end representing more connected thematic clusters of concepts. In other words, concepts in a red-colored theme are more connected than concepts in a blue-colored theme, thus signifying a strong theme.

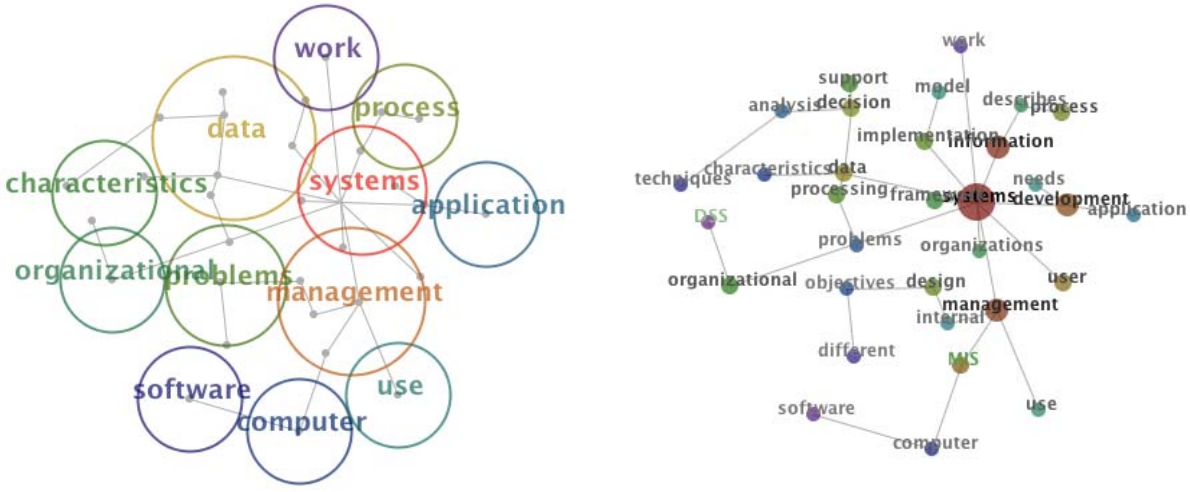


Figure 2: Leximancer example: Concept maps for 1977-1981 Information Systems journal abstracts

The left hand corner of Figure 2 shows a concept map with the concept names suppressed but themes shown. On the right, the same concept map is shown with the concept names shown, but themes suppressed. In the remainder of the paper, we overlay the two representation to show themes and their underlying concepts on one map.

We emphasize here that concept maps *cannot* be interpreted without exploring the context of the underlying data. One of the main advantages of tools like Leximancer is that they provide a consistent, automated approach to the coding of corpi, in a significantly shorter period of time, and with increased repeatability of the coding process, but do not replace the need for the interpretation of the coding. To facilitate the interpretation, however, Leximancer provides a useful interface to the underlying data and its exploration (through facilitating drilling down through the themes, underlying concepts, relevant themes, through to the individual text files where they were identified). The exploration of the underlying data can be done at multiple levels of granularity, starting at the high theme level, and drilling down to individual evidence words placed in the context of the relevant data file being analysed.

Accordingly, we recommend that researchers utilise such tools for reducing bias and expediting the initial coding process, and investing the saved time in an in-depth analysis of the data underlying the resulting coding. In particular, once the coding is obtained, researchers need to make a decision on whether to use the automatically generated Leximancer concept and theme names and support them with explanations, or whether to rename them following the analysis of underlying data (and thus analysis of the context of the concepts and themes). Researchers also have the option to switch off theme generation altogether; however, demarcating themes manually would introduce subjective bias into the coding since theme generation is based on a calculated degree of inter-connectivity of the concepts. Following many such analyses, our approach is to retain the Leximancer discovered concept and theme names, supporting them with an explanation of their context when required. Whichever approach is adopted, researchers need to document their choice of naming approach and the process used for the interpretation of the concepts and themes.

In our work, we adopt the Leximancer generated theme names and explain them based on the Leximancer-supported exploration of a theme's underlying concepts, the concepts' evidence word sets (i.e. the actual words that occur in the text and, together, make up the concept), and sample data supporting the co-occurrence and context of the concepts. To do so, we conduct a drill-down analysis of each concept within a theme and extract supporting quotes to justify the interpretation. In the case of the "systems" theme, for example, using Leximancer to explore the data relevant to the theme indicates that the theme's concepts (systems, information, development, organization, framework, needs)[4] are frequently co-occurring

[4] Listed in descending order of concept strength

through their evidence words (for example, the evidence words for system are: system, systems, evaluation, computer-based, and capabilities, to name a few). The relationships between these concepts become clear when the underlying data is navigated for evidence, e.g. (highlighting added):

"*Systems development and implementation are traditionally approached as a process of designing and fitting a tool to its user's needs. Cases arise, however, due to economic constraints or the desire to standardize, in which an organization directs its subunits to implement a system 'as is'"* (Gremillion, 1980).

Taking multiple instances of such evidence allows us to understand that the theme named as "system" is more accurately referred to as 'systems development and implementation'. The exploration of the underlying data is expedited through the software but must be carried out for each theme and each concept.

Application of LSA and Leximancer

While LSA and Leximancer have been used in a number of academic disciplines (e.g., Hepworth & Paxton, 2007; Landauer, 2007; McKenna & Waddell, 2007), they have only infrequently been applied in Information Systems to date (but never in complement until now). LSA was used in prior research to analyse 14,510 abstracts from 65 IS journals to reveal the intellectual communities comprising the landscape of IS (Larsen et al., 2008). In another study, LSA results from 1,615 abstracts were factor analysed and the resultant topic lists were interpreted and mapped into an existing nomological net (Sidorova et al., 2008). Other studies include multi-lingual document clustering (Wei et al., 2008) and email classification (Coussement & van den Poel, 2008). Focusing on three Information Systems journals specifically, Sidorova *et al*. (2008) hints at an increase in conceptual integration. Hovorka *et al.* (2009) examine the semantic relationships of IS to other business disciplines and suggest conceptual changes among highly ranked journals from each discipline.

Despite the potential to analyze formerly problematic data sets for IS researchers, few examples of IS research exist that use Leximancer for the analysis of text corpus. Stockwell et al. (2009) briefly describe three cases of information processing. Davies *et al*. (2006) examine responses from an open-ended practitioner survey, to examine key factors related to the use of conceptual modeling in IS practice. Martin and Rice (2007) use Leximancer to examine business reports and corporate data to identify themes relevant to risk management. Leximancer has also been used to examine published literature for prevalence of the Design

Science paradigm (Indulska & Recker, 2008) using a seeded list of concepts rather than an automatically discovered list of concepts. These examples indicate the applicability of Leximancer to the examination of unstructured data in IS, yet they have largely focused on *static* 'snapshots' of data rather than analysis of changes over time.

To better understand the relative strengths of LSA versus data mining on the basis of Leximancer, we contrast (Table 1) the two techniques as they were used in this research across a number of dimensions *viz*., approach, application area, size, sample size requirements, suitable analysis, results and outcomes, and limitations. Table 1

Criteria	**LSA**	**Leximancer**
Approach	Calculation of centroids based on vector positioning of text units in semantic spaces	Automated determination of core concepts within the text, based on co-occurrence of evidence words within units of text
Application area	Corpus of text, e.g., conversational data, literature, and other textual documents	
Sample size requirements	There are no minimum size requirements for the size of the text corpi, however, both LSA and Leximancer have been developed specifically to deal with large bodies of text, or large collections of texts. Interpretation of LSA results for sentence length text units is questionable due to the lack of context for meaning. Leximancer is an asymmetric method, thereby, relative sample size of the text corpus does not affect the outcome.	
Application in this research: Trend analysis	Predominantly "what": • What are content centroids and their degree of similarity/ dissimilarity? • What is the trend of such centroids over time and in relation to one another?	Predominantly "how": • How are content concepts composed and how does content change over time? • How are concepts positioned in blocks of textual data and how are they co-occurring with other concepts?
Results and outcomes	• Unique quantification of text units, which can be subjected to statistical techniques (clustering, factor analysis, classification) • Term/document lists	• Co-occurrence matrix of concepts • Concept lists detailing composition of evidence words and their co-occurrence • Visual map displaying themes and concepts and their relative distance to another. Word lists and concept maps can be used for further examination.

Limitations	• Does not incorporate sentence structure • Susceptible to variation in parameters (selection of stop words, weighting, stemming, dimensionality reduction)	Susceptible to variation in size of text block parameter.

Table 1: Comparison of LSA and Leximancer

Research Approach

While our primary focus is on exploring the complementarities of the two analysis techniques through an exemplification on a sample data set, we took care to select a sample of journal data that would be relevant to the IS community. We use insights from prior research (Hovorka et al., 2009) regarding relative changes in conceptual focus as the basis for our exploration of how the two chosen approaches can facilitate the investigation of conceptual changes in published literature over time.

In selecting journals, we used Trieschmann *et al*.'s (2000) determination of a warranted set of top ranked journals. Although Trieschmann *et al*. (2000) identified eight distinct disciplines, we chose to restrict our longitudinal study to three business disciplines due to space limitations and also due to the relevance of the selected three disciplines to readers of the European Journal of Information Systems. In addition to Information Systems, we chose the disciplines of Management and Accounting (see Trieschmann *et al*.'s (2000) journal lists in Table 2) for three reasons. First, Management and Accounting have long been recognized as important reference disciplines to Information Systems (Baskerville & Myers, 2002). Second, Management has been identified as one discipline that is, over time, converging with IS, and Accounting as a discipline that is moving away from IS (Hovorka et al., 2009). Both movements are interesting to study in more detail. Last, a sample of this set of journals is relevant and of interest to the readers of the European Journal of Information Systems, thus providing a captivating and familiar example. We emphasize, however, that the selection of journals shown in Table 2 is not intended to represent the selected disciplines as a whole. Rather we consider them to be samples of important outlets of the disciplines, and use them to demonstrate the capabilities of the quantitative techniques to identify and explore conceptual trends over time. The focus of this paper, therefore, is primarily on demonstrating the capabilities of the techniques and secondarily on examining an interesting phenomenon.

We collected in total 8,544 abstracts from the journals listed in Table 2. While the analysis was carried out in 2008-2009, we selected a 25-year period of data from 1977 to 2001. We selected this period because MIS Quarterly, chosen as one representative journal for the IS discipline, was incepted in 1977. The selected overall timeframe of 25 years for our analysis, while being a sample of convenience, is similar to related studies (e.g., Baskerville & Myers, 2009) and is considered to be of sufficient scope so as to exemplify the application and complementarity of the two techniques. The goal was to examine compare the results or two analytic technique not to investigate the composition of recently published articles.

Fields and Journals	Years	No of Abstracts
Information Systems		
Information Systems Research [ISR]	1990-2001	258
MIS Quarterly [MISQ]	1977-2001	607
Management		
Academy of Management Journal [AMJ]	1977-2001	1938
Academy of Management Review [AMR]	1977-2001	1236
Administrative Science Quarterly [ASQ]	1977-2001	736
Strategic Management Journal [SMJ]	1980-2001	1167
Accounting		
Accounting Review [AR]	1977-2001	1200
Journal of Accounting & Economics [JAE]	1977-2001	442
Journal of Accounting Research [JAR]	1977-2001	960
Total	**1977-2001**	**8544**

Table 2: Core Journals for Selected Academic Business School Fields. From (Trieschmann et al., 2000)

All abstracts were obtained from ProQuest, the online interface of the ABI/INFORM Complete™ (ABII) scientific database. The ABII database is considered the most complete business publications database currently available (Rüling, 2005) and has been used previously in the IS (e.g., Baskerville & Myers, 2009) and broader business context (e.g., Abrahamson, 1996; Rüling, 2005). These characteristics render ABII a suitable sample data source in our study. The abstracts collected from ABII were stored in three databases, one for each of the three disciplines considered.

Following the collection of the required data, we undertake a series of analyses using the two complementary computational techniques. First we examine the semantic relationships

between topic areas published in the three selected data sets (Information Systems, Management, and Accounting journal paper abstracts) over time. This is motivated by recent research that questions the general belief that the business disciplines are conceptually distinct knowledge silos (e.g., Sidorova et al., 2008), and whether computational techniques can reveal patterns of conceptual convergence or divergence in publications. The second question is motivated by the need to identify the conceptual research areas within the data, and how changes in research focus within the selected journal paper abstracts are producing the observed convergent or divergent patterns. Accordingly, we follow a multi-step approach in which we apply Latent Semantic Analysis to determine if conceptual convergence exists, and data mining (Leximancer specifically) to identify related reasons. In doing so, we provide a demonstration of the complementary application of quantitative analyses of textual data.

Because we first wish to ascertain if any convergence or divergence exists, in the **first** step, we perform Latent Semantic Analysis on each of the annual aggregated set of abstracts from the journals. This analysis uncovers longitudinal patterns between the selected journal samples in the three disciplines. We first created a semantic space using all of the abstracts in the sample. The abstracts of the sampled journals from each of the three disciplines were then aggregated by discipline. A sliding window centered on the target year was used to select abstracts from each group for analysis to reduce the effects of short-term fads or special issues. To visualize the results, we chose to hold one of the disciplines constant (x-axis) due to the limitations of representing centroids in a two-dimensional semantic space. This approach allowed the relative convergence or divergence of the other journal centroids to the selected journal centroid to become apparent.

LSA is suited for analysis of similarity between text units both between groups (e.g. disciplines) and longitudinally within groups for the detection of differences or changes in meaning. Term lists of the words that best represent each centroid can be produced and multiple types of statistical analysis can be applied to the LSA output. However, in this study, we are specifically interested in determining which topics in each set of abstracts are producing the changes in the semantic relationships within and between the three sets of abstracts. The LSA similarity analysis does not provide the details of the conceptual drift. Although we could have used human-based interpretation of the term lists (as, for instance, applied by Sidorova et al., 2008), we instead chose an additional computational text mining approach to determine the specific concepts underlying the observed drift.

Accordingly, in the **second** step, we use a series of Leximancer analyses to uncover the most prominent content topic areas, and use these analyses as the basis to explore conceptual drift over time. We perform a series of five Leximancer analyses for each of the three data samples to cover the 25-year timeframe for the Information Systems, Management, and Accounting sample data sets.

The general approach to conducting and interpreting a Leximancer analysis follows the high level process of:

(1) Load the relevant set of data
(2) Run the analysis
(3) Explore core themes (through color coding and theme centrality indicated on the concept map),
(4) Explore relationships and proximity of themes to understand co-occurrence of the themes in the text,
(5) Drill down to identified concepts that underlie the themes – explore their relative distance to understand connectivity and thus context of use in combination,
(6) Drill down to identify evidence words for each concept to understand the exact context of the concept, and
(7) Drill down to explore relevant quotes that exemplify the concepts and provide further context.

We note that this application process is similar to that performed in exploratory factor analysis (Gorsuch, 1997), where a large set of (quantitative) variables is explored to uncover underlying factor structures that can be used to reduce the set of variables to a lower number of unobserved variables called factors. The analogy to a Leximancer analysis is that themes (factors) are identified in a corpus of text based on the co-occurrences (loadings) of key underlying concept terms (measurement variables).

Performing the above described analysis for each of the 5-year data sets within each of the three selected sets of journals provides us with a collection of valuable automatic outputs that can be used to identify, reason about, and further explore, core concepts and themes within the data. The analysis also facilitates the efficient navigation of the underlying data, which helps in the interpretation of the themes and concepts. In particular, Leximancer creates for each analysis a fully interactive concept map together with a series of theme connectivity lists for each map. We utilize the maps to reason about the core topical areas of discussion in the relevant data set, and also take into consideration the theme connectivity lists (exemplified in

the next section) to reason about the breadth of discussion topics within a particular data set. In particular, we look for situations where the number of strong themes in a given data set differs strongly from another. While the variety of themes across data sets has an obvious implication about the changing focus of discussion, an indication of how many strong themes exist within a data set provides a useful indication of breadth. For example, if one data set consists of only three strong themes and another consists of 10 strong themes, then we can draw the conclusion that the first data set is more focused on three core topics while the other is diversified.

Having performed the second step of analysis, we obtain an in-depth understanding of the themes in each 5-year period and an indication of the changing breadth of discussion across the data sets. In interpreting these Leximancer outputs, which includes navigating the underlying data to understand the context of the themes, we can then move on to the identification of conceptual drift within each data set.

Accordingly, in a **third** step, we move forward to the exploration of the strongest themes in each time period, the exploration of data underlying those themes, and the identification of thematic overlap between data sets. This step relies on the researcher to use the Leximancer concept map to understand the context of each strong theme. Using Leximancer to drill down into the underlying data to interpret the strongest theme in each 5-year data set, we can construct drift maps of the core concepts emerging from the sampled journal abstracts over time.[5] These drift maps (exemplified in the following section) allow us to visualize the conceptual drift longitudinally *within* a set of journal abstracts from one domain (e.g. Information Systems), and are central to our understanding of how the core foci have changed over time. However, the drift maps do not identify overlap *across* domain data sets since the overlap is likely to be in the form of increased discussion of same concepts, rather than identical themes within the same context. Leximancer is unable to automatically perform an overlap analysis of the themes identified across the data sets – which is what is required to identify which topics are responsible for any convergence (or divergence). Such an analysis can be supported by Leximancer through its data browsing features in the interactive concept map, but still requires major manual intervention.

[5] We focus on one theme (the strongest theme) in our analyses as a means of exemplification of the techniques, but researchers can determine an appropriate number of themes to look at.

Accordingly, to determine if overlaps exists between the themes discovered in the 5-year Leximancer analyses, in the **fourth** step, we conduct a Leximancer-supported process of navigating and interpreting the maps from step two to judge the level of actual overlap between themes.

In the first instance, we consider the concept maps created in our Leximancer analyses (step two) to identify, for each of the timeframes, the amount of thematic overlap with same timeframe concept maps for other datasets. An increase or decrease in core theme overlap, would, respectively, suggest the convergence or divergence of the datasets. This step also allows us to reason about the direction of the relative thematic convergence or divergence seen in LSA. This process has some limitations relating to subtle changes in the theme context (through the change of underlying evidence words, again highlighting the need to drill down into the concepts to understand their context).

We also note that convergence or divergence of journal publications ascribed to different disciplines may in some instances occur without a significant core thematic overlap. This situation would suggest that, without overtly changing the main themes of discussion, one disciplinary journal may increasingly discuss concepts that are central to a journal associated with another discipline, thereby resulting in a subtle change of theme context (i.e., a change in underlying concepts, or, even more subtly, a change in the underlying concepts' evidence words). This limitation is common to our approach where the journal abstract sets from different disciplines are analyzed in Leximancer in isolation and the relationships between concepts in these disparate data sets is thus not explored and not taken into consideration during the concept learning phase of the Leximancer algorithm. It is possible to supplement such analyses with a second Leximancer analysis performed on the two data sets jointly. Leximancer, in particular, allows for such an analysis by means of *file concepts*, which can be used to explore the relationships of concepts within each of the two datasets within a single combined analysis. Thus, in a final step, we conduct such an analysis to gain a better understanding of the core concepts that are common to a joint data set. We note that of particular interest is the delta-analysis of concepts that are core to a data set of a particular domain versus the concepts that are core in an analysis of a combined data set.

Analysis and Results

Step One –Latent Semantic Analysis

In this research, LSA was used to calculate the centroid for the aggregated abstracts from each discipline-specific journal set. The centroid is the point in semantic space calculated from the term relationships in each set of abstracts. The angle between the centroids for each sample (e.g. the angle between the centroid for the IS abstracts and the centroid for the Management abstracts) was plotted using a sliding window protocol to visualize the convergence or divergence over time of the concepts contained in the aggregated abstracts. A decreasing angle between the centroids of each group of abstracts indicates decreasing distance in semantic space and therefore greater similarity in topics. The opposite also holds true. Figure 3 displays the results graphically.

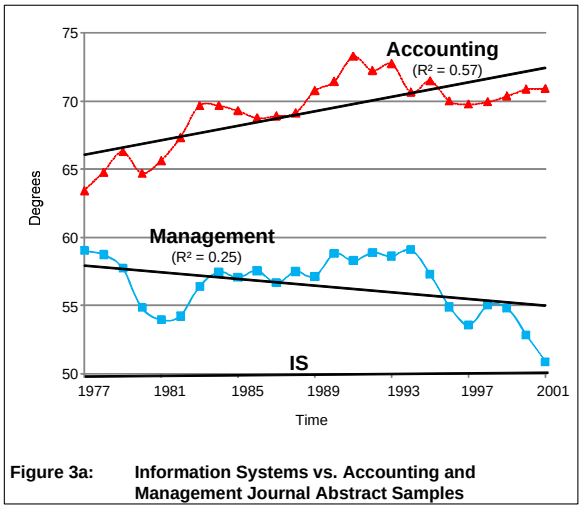


Figure 3a: **Information Systems vs. Accounting and Management Journal Abstract Samples**

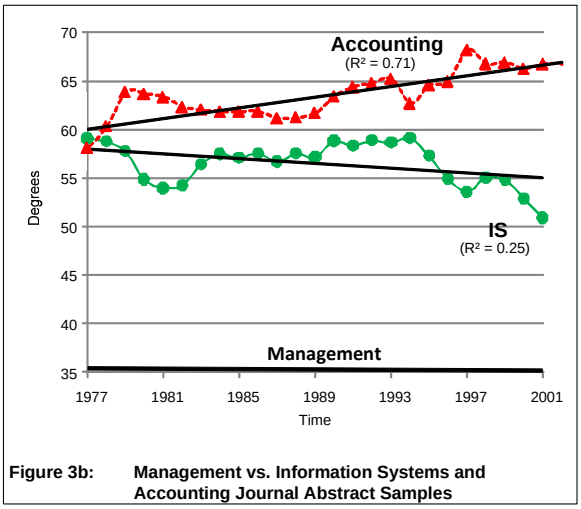


Figure 3b: **Management vs. Information Systems and Accounting Journal Abstract Samples**

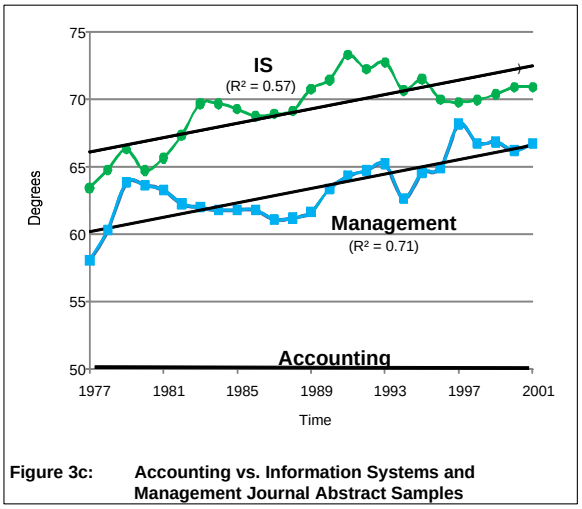


Figure 3c: **Accounting vs. Information Systems and Management Journal Abstract Samples**

Figure 3: Conceptual Trend Analysis of Three Selected Journal Abstract Samples

Perusal of Figure 3a shows that the angle between the centroid for the Management journal abstracts sample, and the centroid for the Information Systems journal abstracts sample is decreasing with time. This finding indicates that the research topics in the two sets of journals are becoming more semantically similar even if they are expressed in different words. At the

same time, the centroids for the Accounting journal abstracts sample are diverging from those of the Information Systems journal abstracts sample, indicating that concepts are becoming less similar over time. A significance test shows this slope trend to be significantly different from zero. This pattern is consistent across Figures 3b and 3c, in which each of the other disciplinary journal abstract samples are held constant as the baseline. For example, Figure 3c shows that the centroids for Management and IS journal abstract samples, are each diverging from the abstracts in the Accounting sample.

Step Two – data mining analysis

We exemplify here the Leximancer analysis of the trends between the Management and IS journal abstract samples only, while the Leximancer analysis of Accounting journal abstract samples is not reported here (but follows the same analysis process and reasoning).[6] This analysis provides us with a set of concept maps and theme identifications. In addition to understanding which concepts belong to a theme and what evidence words make up a concept, to facilitate the exploration of conceptual drift using Leximancer, one must also understand theme strengths. In addition to the concept map, Leximancer analysis provides a list of relative theme strengths. These lists can be interpreted, with the help of underlying data exploration, as indicating when and how strongly evidence words are related to a specific concept in the related visual map.

We carried out this analysis and exploration for all time periods in the data. The analysis allowed us to generate concept maps and theme relevance distributions and navigate these to understand their context. Figure 4a-e display the core concepts and the themes in the abstracts from the selected IS journals, and their co-occurrence in relative position to each other, in a series of five maps, one for each 5-year interval considered. Additionally, Figure 5 shows the most frequently occurring themes, and their relative strength, in descending order, for each of the 5-year intervals. The details for the Management journal abstract sample are presented in Appendices B and C.

[6] The analysis is available from the authors upon request.

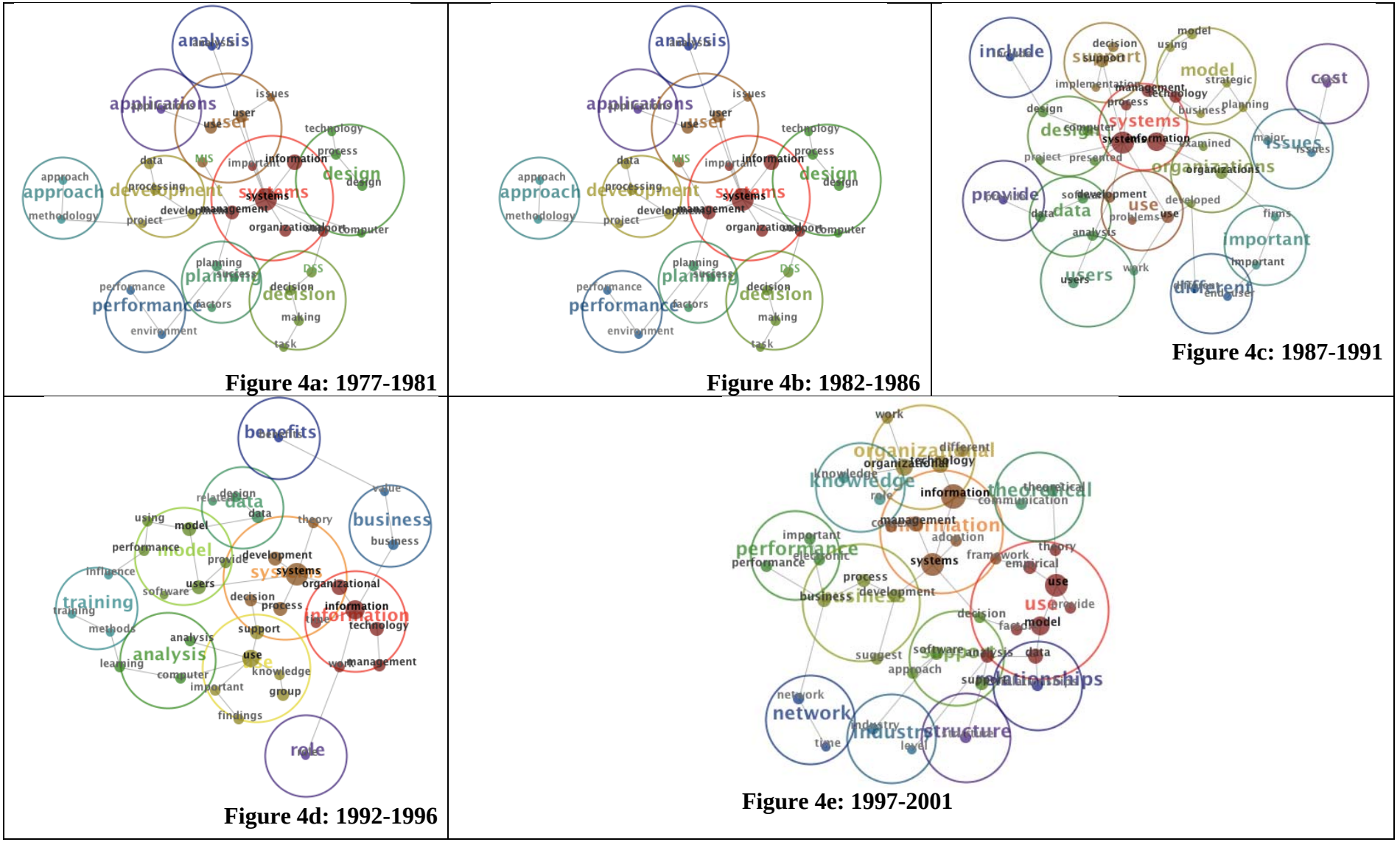


Figure 4a: 1977-1981

Figure 4b: 1982-1986

Figure 4c: 1987-1991

Figure 4d: 1992-1996

Figure 4e: 1997-2001

Figure 4: Concept Maps for the set of Information Systems journal abstracts, in 5-year intervals

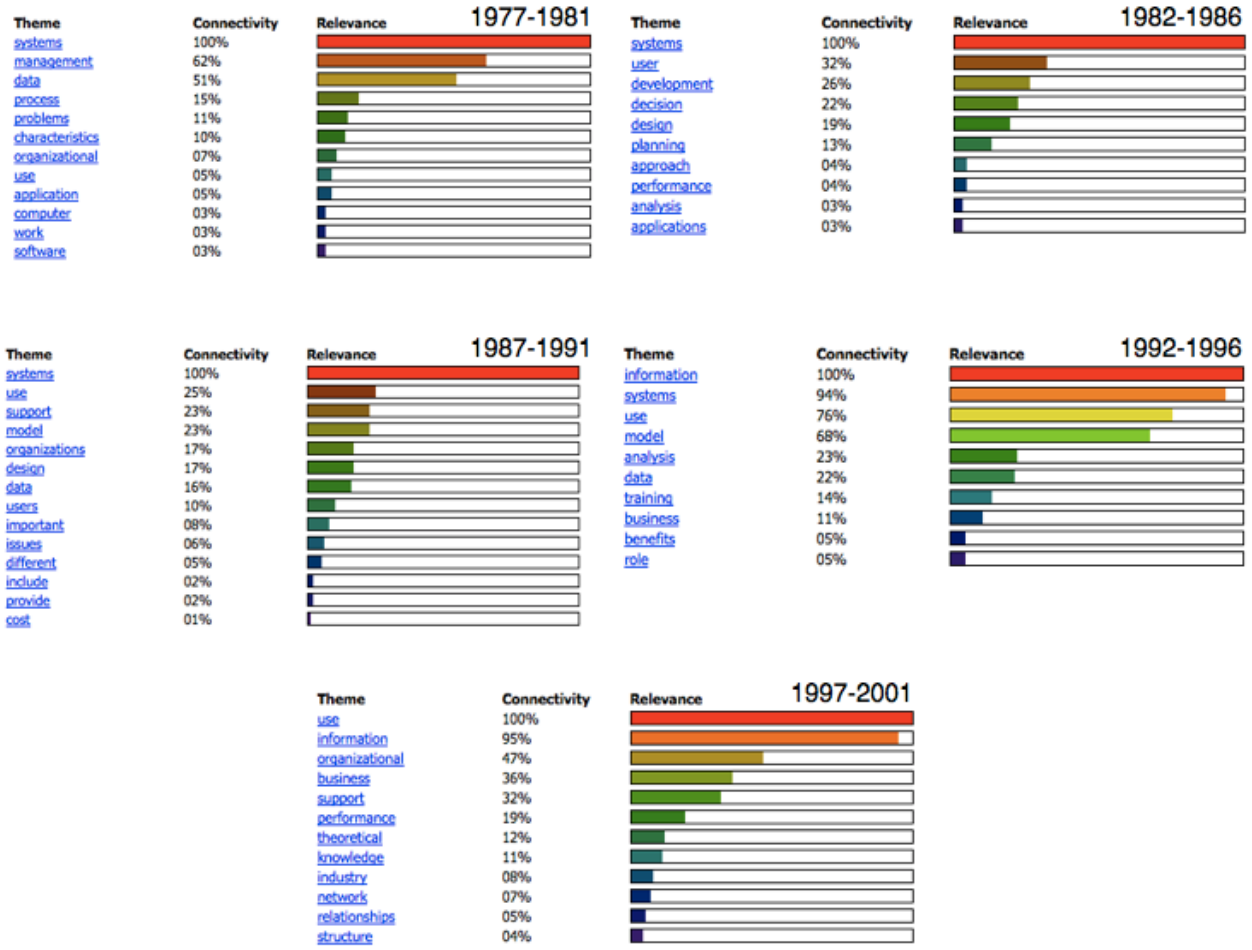


Figure 5: Theme Connectivities and Strengths in the set of Information Systems journal abstracts, in 5-year intervals

Based on Figure 4 and Figure 5 we offer an explanation of the core themes uncovered through the series of Leximancer analyses.

In 1977-1981, the emerging concepts (see Figure 4a) and the relative theme strength distribution can be interpreted, with the help of underlying data exploration, as indicating that whenever evidence words related to the concepts of the management theme are mentioned in the text, there is a 62% co-occurrence rate with evidence words related to the concepts of the *systems* theme. Likewise, whenever evidence words related to the concepts of the *use* theme are mentioned in the text, there is only a 5% co-occurrence rate with evidence words related to the concept of the *systems* theme. The *use* theme exists due to the existence of a strong *use* concept (which is made up of the evidence words of: use, problem solving, choose, but which excludes *user*, which is a concept that relates *use* to *systems*). Exploration of the underlying data provides an indication as to the context: "The authors advocate providing or at least simulating *user* capabilities early in the *systems development* process. Such an approach is made possible by the *use* of an online relational-type Database Management *System*" (Berrisford, 1979).

Perusal of the data on relative theme strength distribution in 1977-1981 further indicates a distribution of central themes in the selected Information Systems journals that is similar to a power law distribution – i.e. we see that only few themes are highly connected with the rest being towards the tail end of the distribution. This indicates a concentrated focus on a relatively narrow set of themes. Exploration of the data, and further exploration of the core concepts within the shown themes in Figure 4a, indicate that the systems theme focuses most on system development, while the use theme is strongly focused on the use of systems and development approaches.

In the set of Information Systems journal abstracts in 1982-1986, Leximancer analysis and subsequent data exploration, indicates that the strongest themes of "systems" "user" and "development" (see Figure 4b) relate commonly to discussions of systems development, the involvement of users in the development process, and, to a lesser extent, the use of decision support systems (in relation with the weaker "decision" theme). These linkages are confirmed through exploration of the underlying text, for example:

"The authors argue that effective management of DSS development requires: a) An explicit plan for the full development life cycle; b) Careful assignment of responsibility for DSS

development; c) Appropriate user involvement and direction; and d) On-going user needs assessment and problem diagnosis." (Meador, 1984).

For the map shown in Figure 4b, the related distribution of theme strengths, shown in the top center of Figure 5, indicates that the theme of "systems", and to some extent the themes of "management" and "data" are the most connected and relevant themes within the data set, implying that the concepts underlying those themes are frequent and frequently occurring together through their underlying evidence words.

In the data set representing 1987-1991, we identify a strong systems theme that is comprised of concepts relating to systems, implementation, development, information, and technology (see Figure 4c). The main focus in this time period is still on systems development but we note a change in vocabulary around this time, when the prior use of the term "systems" or "management systems" starts to be more frequently replaced with the term "information systems". The thematic focus of this data set is confirmed through the exploration of underlying data, and exemplified through, e.g.:

"Bull HN Worldwide Information Systems planned and implemented an expert system to be used for troubleshooting the maintenance of its page printing system. A case study identified important aspects of the design and development of the system" (Braden, 1989).

In the data set representing 1992-1996, our analysis indicates that the themes of "information", "systems" and "use" are the strongest themes (see Figure 4d). The context of "systems" still includes concepts of development, but now additionally has a strengthened focus on decision support. Note, as in the previous time period, a tendency to refer to systems as "information systems". The main focus of this dataset is the use of information systems. This is exemplified by quotes from underlying data and can, optionally, additionally be confirmed with Leximancer path analysis that shows the strongest path between any two concepts.

Finally, Figure 4e displays the data set representing 1997-2001. We see a notable change in that systems and information have a stronger and more frequent co-occurrence, so much so that they now are concepts within the same theme ("information"). This situation also implies that evidence words relating to the "information" concept occur more frequently than those relating to the "systems" concept. Our exploration of the underlying data confirms an additional focus on information (as opposed to mainly the information systems from which the information is extracted). In this time period the strongest themes in this data set are those

of “use”, “information” (which includes the “system” concept), “organizational” and “business”. The use of information is a core focus in this data set. Further exploration of the underlying data indicates that the “organizational” theme relates mainly to organizational knowledge and organizational roles, and to a lesser extent performance, while the “business” theme relates mainly to performance of the organization. The focus can be exemplified through, e.g.:

“In this study, we fill this gap by comparing four newly-industrialized economies (NIEs) with regard to the impact of IT capital on business performance.” (Tam, 1998)

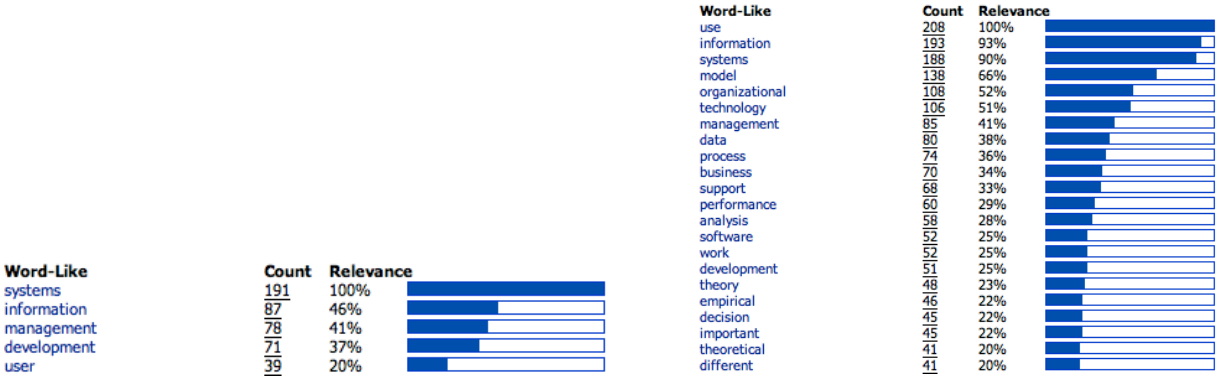
or

“For the information industry to bridge rather than divide further the global economy to information rich and information poor, we need to understand how firms, particularly local firms, can pioneer or participate in the information industry in emerging economies that do not inherently embrace information as a valued business resource” (Jarvenpaa & Leidner, 1998).

Turning back to the graphs for theme connectivity and strength in Figure 5, we note a power law-like distribution of strongest themes in the years 1977-1991, dominated by the centrality of the “systems” theme in this timeframe. We further note that, after 1991, this power law-like theme distribution is being replaced by a distribution of terms that violates power law or Pareto principles. For instance, during 1997-2001 we find that five themes feature a connectivity of over 30%. We interpret this finding as an important sign for increased diversity in the themes of articles published in the selected journals, over recent years. In other words, we see a widening of the focus of discussion since 1992.

Indeed, drilling down to a concept level in the Leximancer analysis identifies a trend in the broadening of core concepts under discussion. Figure 6 shows the relevant list of core concepts within the Information Systems dataset (concepts 20% or more connected to other concepts are displayed only) in 1977-1981 and 1997-2001 respectively. The graph on the left-hand side of Figure 6 indicates that the discussion of core concepts is strongly focused on systems development, with some aspects of user issues (which become more prominent in 1982-1986 – refer to Figure 4a,b). The right-hand graph in Figure 6, however, shows a strong increase in the number of core concepts emergent from the Information Systems dataset (from five to twenty-two) by 2001, providing evidence that, and which, additional topics

have been more significantly embraced in the Information Systems dataset relative to 1977-1981.



1977-1981 **1997-2001**

Figure 6: Information Systems dataset concepts in 1977-1981 and 1997-2001 (with 20% or more likelihood of discussion)

Step Three – conceptual drift analysis

Following the exemplified analysis of the Information Systems dataset, we move forward to the identification of *conceptual drift*. To carry out this analysis, we utilize the Leximancer interface to drill down on the main themes and their underlying concepts (and their evidence words) to gain an understanding of the context of the themes. We present this analysis in the following sub-sections, structured per journal discipline area.

- Information Systems

Analysis of the Information Systems journal abstract dataset uncovers that the selected journals have undergone a number of shifts in research focus. In the early years of the considered dataset, the themes of published research centre around systems development. Indeed, in the first three considered 5-year periods (1977-1991), the focus of the journals was on systems development. Within that timeframe, however, the aspects of the system that were the focus of study have shifted. Our analysis shows that the initial focus was on systems development, which then extended to systems development with consideration of user issues. In the next time period (1992-1996) we observe a shift in focus away from systems development and towards the use of information systems. Note that this situation does not indicate that research on systems development no longer takes place, but, rather, that it has been outweighed by other themes as a central focus of the journal abstracts in the sample. We note a further shift towards the use of information in organizations (in 1997-2001 dataset).

We conceptualize the shift of strongest themes emerging from the data considered, in 5-year intervals, in the conceptual drift map shown in Figure 7. We also note that the analysis identifies a shift in vocabulary from "systems" or "management systems" to "information systems" over the years.

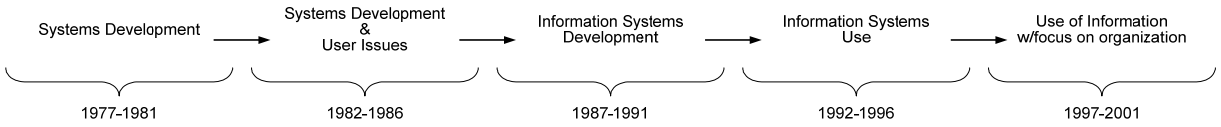


Figure 7: Conceptual Drift of Dominant Themes in the selected sample of Information Systems journal abstracts, in 5-year intervals

- Management

Similar to the analysis of the selected Information Systems journal abstracts, we subject all selected Management journal abstracts from our data set to a 5-year interval analysis. Leximancer analysis (see Appendix B and C) again facilitates the identification of the movement of strong themes over time, which we capture in a map of the conceptual drift in the journal abstracts sample considered (Figure 8).[7]

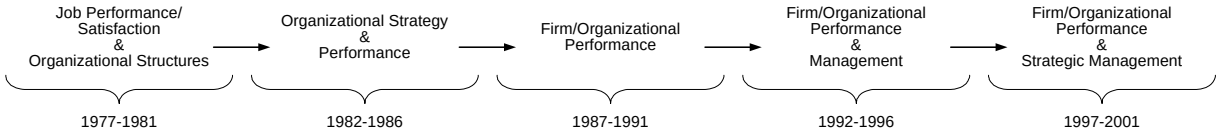


Figure 8: Conceptual Drift of Dominant Themes in the selected sample of Management journal abstracts, in 5-year intervals

For the selected Management journal abstracts, we note that the main themes of research have evolved from those focused on organizational structures to a sustained focus on organizational performance over more recent years. We see a transformation in the topics, from individual performance, to organizational strategy and organizational performance. Yet, since 1987, the core research theme reflected in the journal abstracts sample remains on the performance of the firm, with lesser foci also on strategic management and process management.

When compared to the Information Systems journal abstracts sample, the set of Management journal abstracts has a relatively consistent research focus over the last 15 years; being most

[7] Exploration of the context of the dominant themes follows the same process as in the case of Information Systems journal abstracts, and is available from the authors upon request.

concerned with performance, and the impact of various types of management approaches (organizational/strategic/process) on organizational performance. We interpret this finding as sign that the Information Systems journal abstracts were in the early years focused on defining development approaches and aspects related to IS development, and, over the years, as IS matured, had an increasing focus on the use of systems and the use of information for business purposes – a finding relevant to the discussion about the ongoing diversity debate in IS (Vessey et al., 2002). This finding also provides some insight into why the Information Systems journal abstracts and the Management journal abstracts are indicated by LSA to be converging.

Step Four– Identification of Common Themes

Identification of any thematic overlap requires a careful analysis of context of each theme before a judgment of overlap can be made. In carrying out this analysis, for each theme that had the potential for overlap (i.e., due to same or similar naming), we again performed a drill down analysis to understand whether the contexts of the two themes in question (and their relevant concepts) were sufficiently similar. While the drill down process is an interactive one and the length of the paper precludes us from presenting the whole justification of analysis for the exemplary application of Leximancer, we support the analysis with direct quotes from the underlying data to elucidate the similarity of context.

Considering the Information System and Management journal abstract samples, we observe limited thematic overlap between the two in 1977-1981 (refer to Figure 4a and Appendix B, respectively). The only two close themes are those of “management” (Information Systems) and “managers” (Management), and “organizational” (Information Systems) and “organizations” (Management). However, using Leximancer to drill down into the concepts underlying these themes we uncover that “management”, in Information Systems journals, is closely related to systems, whereas “managers”, in Management journals, is related to strategic planning and business strategies. Accordingly, we interpret the two themes as being dissimilar. We observe, however, some convergence as a result of the “organizational” and “organizations” theme. The “organizations” theme in the Management dataset includes the concepts of “organization”, “organizational”, “model” and “behavior” and there is observed co-occurrence of the “organizational” concept with a “framework” concept. In the IS dataset, the “organizational” theme is based on concepts of “decisions”, “characteristics”, “analysis” and “support” but the “organizational” concept also co-occurs with a “framework” concept, thus signifying some overlap.

A drill down analysis uncovers a range of quotes that indicate the above analysis. For example:

"A method is set forth to explore systematically how Management Information Systems (MIS) influence the horizontal distribution of power among departments and interdepartmental communications. MIS is defined as a computer-based system used for providing information to support operations, management, and decision making. A framework for examining MIS and evaluating them is given. Interdepartmental communications increase with the use of MIS. The real or perceived power of organizational subunits is assessed considering: 1. coping with uncertainty, 2. nonsubstitutability, and 3. pervasiveness." (Saunders, 1981)

In 1982-1986, we note the existence of one common theme – that of performance, which in both cases relates to organizational performance. While this theme is a weak theme in the Information Systems dataset (4% connectivity, as per Figure 5), it is nonetheless an overlap that indicates a closer relationship between the two datasets. The finding appears in-line with the trend of the LSA results shown in Figure 3b. Investigation of further timeframes identifies no core thematic overlaps in 1987-1991 and 1992-1996, indicating a divergence relative to the 1982-1986 timeframe. Cross checking with the LSA findings, we again observe a slight divergence at this point in time. Drilling down into the data with Leximancer, provides an insight for the closer relationship of the two datasets, with Information Systems publications considering duties of the Information Systems manager in changing environments. For example:

"The work presented in this article relates directly to perhaps the most serious problem facing the Information Systems manager in a large, complex organization today, namely how to plan and manage in a rapidly changing, high-demand, resource-limited environment The article describes an organizational change effort undertaken within a major data processing organization to seek improvements in four broad areas data. center production performance, responsiveness of the systems development activity, management control and decision making, and long range and operational planning processes." (Loftin & Moosbrukker, 1982)

Yet, in the last timeframe of 1997-2001, we identify three core theme overlaps – those of "information", "organizational", and "performance", with the first two being highly relevant themes in Information Systems (95% and 47% connectivity, respectively), and

"performance" having increased to a 19% connectivity and higher relevance than in 1982-1986. For the Management dataset, the themes of "organizational" and "performance" are among the most relevant themes (both with 84% connectivity) and "information" being a weak albeit existing theme (at 6% connectivity). Contrasting this finding to the relative movement of the centroids of the journal abstracts in Figure 3, we can speculate that the stronger observed convergent conceptual drift of the IS and Management journal abstract samples is due to the increased consideration of mutually relevant topics since 1997. More specifically, it appears that, since 1997, both Management and IS journals have uncovered an interest in research pertaining to the "information", "organizational", and "performance" themes, with Information Systems appearing to adopt a stronger organizational and performance focus, and Management, in turn, increasing focus on information use, suggesting that both the two domains are, partly, responsible for the noted convergence. Hence, while over the years the two sets of journal abstracts have been on a converging trend, there is a sharper increase noted since 1997 (as indicated by LSA analysis in Figure 3). Similarly to the 1982-1986 analysis, a drill down in Leximancer provides the researcher with an indication as to why such overlap is present. For example:

"As organizations implement more and more distributed work arrangements such as telecommuting, there is a need to understand the determinants of success of this new work setting. This research investigated three variables believed to impact outcomes in telecommuting: the availability of information system technology, the availability of communication technologies, and the communication patterns of telecommuters within their work groups. Two perspectives are used in this study. The direct effects of these three variables on perceived productivity, performance, and satisfaction were tested." (Belanger et al., 2001)

As noted in the Research Approach section, manual overlap analysis is not always suitable in isolation, since convergence or divergence of text units may in some instances occur without a significant core thematic overlap but rather through gradual and subtle changes in theme context. We thus exemplify how a Leximancer analysis of the two data sets jointly can facilitate further interpretation of the changing topics. In Figure 9, the collective Information Systems dataset is indicated by a "FILE_InformationSystems-" concept, with an appended timeframe. Likewise, the Management dataset is represented by "FILE_management-" concept. Specifically, it shows the relationships of the Information Systems dataset for 1977-1981 and how the data set relates to concepts that were identified to be the strongest common

concepts across the two data sets. Visual inspection indicates that, while there are concepts that are common to both data sets, the strength of the connection is weak (relative to the Management data set, as evidenced through the color of the links in comparison to Management connections in Figure 9).

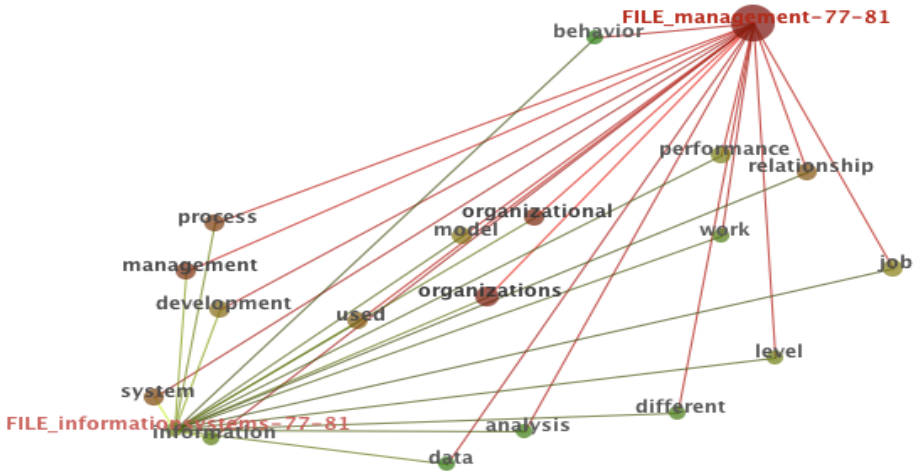


Figure 9: Top 40% in the combined Information Systems and Management 1977-1981 dataset, and their connection to Information Systems and Management publications

In particular we identify a situation that indicates that the Information Systems journal abstracts have some connection to organizational aspects (even though they are not the core themes in the Information Systems dataset alone – see Figure 4) but not a strong connection (indicated by the green connections to such concepts), as opposed to stronger and closer concepts of management systems and development. Interestingly, we note in this concept map that "information" emerges as a strong concept in this analysis but not in the isolated analysis of Information Systems journals for this time period (see Figure 1). Again, this situation occurs because the frequency of the information concept within the Managemet journal abstract dataset in this time period ensures that it is an identified strong concept, and hence, weak links, are identified to the Information Systems dataset.

Another important aspect to note in this joint analysis is the relative placement of concepts to the file concepts. From Figure 9 it is visible that concepts of "performance", "relationship", "work" "organizational" are closer to the sample of Management journal abstracts than the sample of Information set of journal abstracts.

Such comparisons can also be done across time periods for the purposes of comparison. Figure 10, for instance, displays the results from the join analysis of the combined Information Systems and Management 1997-2001 dataset.

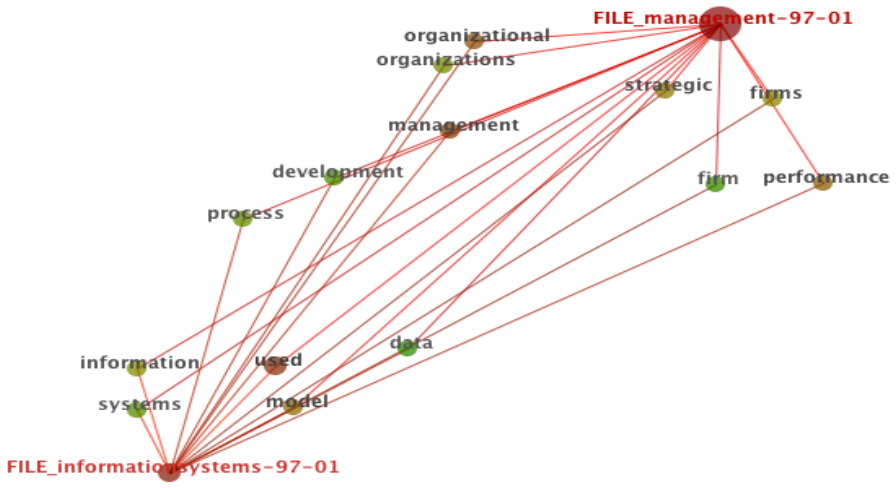


Figure 10: Top 40% in the combined Information Systems and Management 1997-2001 dataset, and their connection to Information Systems and Management publications

Notably, in Figure 10 we note a strengthening of the links between the concepts associated with the Information Systems file concept and those associated with the Management file concept. This is indicated by red coloured links between concepts that in prior years (e.g,. 1977-1981, as per Figure 9) were coloured green (as weak). For instance, we note that the concepts of "performance" and "data", which featured only weak (green) links to information systems in 1977-1981 now have stronger connects (indicated by red lines in Figure 10). This shift over time can be interpreted as indicating that the joint discussion of concepts like "performance" and "data" increased in the journals associated with the two disciplines.

The pairwise comparison between Figure 9 and Figure 10 further suggests that concepts that were closely linked to one discipline in 1977-1981 (e.g., information to information systems, and behavior to management, see Figure 9) can, over time, transcend towards concepts shared by both file concepts. For instance, we note that the concepts "management" and "development" that were positioned closely to the Information Systems file concept in 1977-1981, now appear positioned centrally between the Information Systems file concept and the Management file concept (see Figure 10), indicating a balanced discussion in the respective literature.

Last, we note that such movements towards central positioning between file concepts is also accompanied by concept movements from the center to one file concept only. For instance, the "model" concept that was positioned centrally between Systems file concept and the Management file concept in 1977-1981 is in Figure 10 positioned close to the Information

Systems file concept. This movement can be interpreted as indicating that the discussion of models has disproportionally increased in IS literature relative to Management literature.

As these selected analyses of common themes and their movement indicates, human analysis and interpretation of data sets can be significantly aided through the two techniques we considered. We showed not only how we can identify conceptual drifts in literature but also how we can selectively explore and examine in detail the underlying concepts, their relationships as well as their movements over time.

Implications

We can identify a number of opportunities for future research in this area. First, we have shown how Leximancer, a data mining tool, in complementary use with LSA, can be used to identify the movement of research foci over time, as represented in published literature (the "what"), and also how this conceptual drift is occurring (the "how"). We note the usefulness of LSA to identify conceptual convergence or divergence of separate datasets, and the applicability of Leximancer, as our data mining tool of choice, to uncover the underlying reasons for the evident convergence or divergence trends. Based on this work, we can identify several further avenues for research.

Most notably, scholars could apply LSA and Leximancer together in the analysis of other large textual datasets. In IT risk and security research, for example, the complementary use of the two techniques could be useful for forensics or profiling purposes (e.g., Liang & Xue, 2009), to identify key topics (e.g., those indicating activities related to threats) in conversational data, or to understand the relationships and underlying movement of conversational topics and how they relate to certain conversational outcomes (e.g., subsequent actions undertaken by the groups studied). Research on the use of information technology could use the approach presented to identify positive/negative moods or affect, how they evolve over the course of time, and how they relate to the adoption or use of technology (de Guinea & Markus, 2009). Similarly, research on identity development in virtual spaces (e.g., Zhao et al., 2010) could use our approach to better understand which themes characterize their existence in virtual spaces (e.g., Facebook status updates, Second Life conversations, etc), and how the co-occurrence and composition of these themes contribute to the identity development process.

Second, the type of analysis we exemplify for the selected journal abstracts is relevant and may be extended to conferences, other IS journals (e.g., those included in the AIS Senior Scholars' Basket of Journals) or to journals from other disciplines in which IS concepts are salient (e.g. adoption, use, governance of information systems in medicine, finance, sustainability science). The combination of the analytic techniques we consider offers the opportunity to examine where literature published in various academic disciplines overlap or converge in the conceptual areas that researchers are pursuing. Recognition of inter-disciplinary areas may lead to a greater degree of knowledge integration in academic disciplines. Furthermore, the connection between academic and practitioner literature could be conceptually mapped, thereby potentially providing a clearer picture of the degree of relevance of academic research to practice and visa versa.

Third, the scope of analysis could be focused differently. We examined the type of content topics subjected to research in the different journals we considered. Future analysis could focus on the methodologies used in studies reported in these or other journals, and how methodological choices vary over time. Such research could meaningfully extend our literature on methodological diversity (e.g., Vessey et al., 2002).

Fourth, we identify opportunities on a methodological level to further examine the similarities and differences between quantitative and computational versus manual techniques for the analysis of corpus of text. This research could, for instance, contrast the results obtained through quantitative approaches to those obtained through human coding (via tools such as Nudist or NVivo, for example) in individual or multi-rater coding processes. Such research could draw conclusions about the relative internal and external validity, as well as reliability, of different approaches to text analysis across, for example, various sample sizes or data formats. Also, researchers may wish to vary some of the settings in the use of LSA or Leximancer to explore different ways to examine data sets. For instance, Leximancer also permits *seeded analyses*, which do not rely on automatic concept generation (as used above) but instead use a seed list of concepts that are specifically of relevance to the study at hand. This form of using Leximancer could, for instance, lend assistance to the analysis of fashion waves in IS research, as introduced in Baskerville and Myers (2009). Such an analysis can further assist the identification of knowledge from a body of text, but is generally used in a confirmatory rather than exploratory manner, viz., when the researcher looks for evidence of an a-priori theme or concept (which was not the case in our application of Leximancer). While we do not exemplify this analysis in the paper (it has no added value since we are not

looking for specific concepts), we note the possibility due to its potential application in the Information Systems discipline (e.g. analysis of confirmatory case studies, interviews, etc).

Conclusions

Contributions

Perusing data from over 8,500 published abstracts collected from core journals in three discipline areas, we applied two quantitative computational techniques for the examination of large corpus of text. Specifically, we demonstrated the use of two quantitative techniques for the complementary longitudinal content analysis of large textual data sets that would be difficult, if not impossible, to perform manually. We show how the results from the two analyses can be used together to extend our understanding of *what* dominant high-level themes are in the published literature of selected journals, *how* these themes are composed, and *how* they contribute to the thematic movement of ideas published in the journals over time. The identification of themes around which these journal abstract sample datasets share (or diverge in) interest provides insights into the types for problems and lines of inquiry where interdisciplinary research might be fruitful. This integration of knowledge can lead to recognition of concepts of interest in one discipline that are also important in research of other related disciplines.

Our work illustrates in a 'proof of concept' approach how two modern quantitative computational analyses can be used, in isolation as well as in complementary fashion, to aid the content analysis of a large corpus of text. We show how the results of one analysis (LSA) can be used to inform our understanding of the trends between separate datasets, and we demonstrate how a a text mining analysis (using Leximancer) can be used to provide further insights for the underlying rationale of the outcomes of the LSA analysis. We further demonstrate how the outputs generated by the techniques can aid interactive human exploration and analysis of the data.

We draw our sample of journal abstracts from the Information Systems, Management and Accounting disciplines, so we can highlight how the two approaches *viz.* LSA and Leximancer, can help researchers contribute to the ongoing reflexive discourse (Baskerville & Myers, 2009; Ramiller et al., 2009) on how we, as academic scholars in the Information Systems field, set and pursue our research agendas over time, and in relation to important reference disciplines (Baskerville & Myers, 2002).

Limitations

We consider our work an exploratory study intended to demonstrate the utility of two quantitative content analysis techniques *viz.* LSA and data mining, to better visualize conceptual drifts and thematic topic areas within and between core journals from academic disciplines over time. While all care was taken to increase the rigor and objectivity of this study, we consider limitations relevant to the use of the quantitative computational techniques we employed.

We note that LSA uses a "basket of words" approach that flattens the semantic content of abstracts to a single point (the centroid) and may thus be susceptible to bias pertaining to the categorization of abstracts that contain multiple concepts. It also removes syntax, thereby decreasing context related to the terms. In addition, although Leximancer reveals the existence of concepts underlying a corpus of text, it relies on frequency counts to determine the relative strength. This may lead to the inability to discern concepts that are infrequent, and, thus a weakness of identifying newly emerging themes that are appear infrequently. Moreover, despite the reproducible analysis, the navigation and exploration of data to interpret the Leximancer concept maps, albeit supported in a consistent and transparent manner, introduces a risk of researcher bias.

The first two limitations in particular are inherent with both techniques. Yet, we have shown in this study that the different approaches to data examination may in fact also be used in complementary fashion. The combined application, therefore, allows researchers to not only examine results in more breadth through the different 'lenses' offered by alternative techniques, but also to mitigate limitations pertaining to each technique individually. In this paper, however, we restricted the individual application of each of the two techniques. Therefore, although our proof of concept highlights the ability to identify longitudinal conceptual drift, we point out that both LSA and Leximancer can be applied to a wide variety of other analyses of any type of textual data. For instance, we point the interested reader to the five methodological recommendations pertaining to the use of LSA described by Evangelopoulos et al. (2011).

Furthermore, while we do not consider our identification of conceptual drift in the journal abstracts to be representative of the disciplines, our analysis reveals interesting trends in highly rated journals from each field. While we chose to focus on abstracts to remove the potential bias that the methodology and reference sections would have on the analysis, we

consider it an interesting avenue of study to apply LSA and Leximancer to the methodology sections to see how research approaches have shifted over the years.

Finally, due to our focus on the complementary application of LSA and Leximancer, we note that the individual analyses could have been performed in different ways. Specifically, for illustration purposes, we restricted our LSA and Leximancer analyses, in turn, to standard settings. For example, perusing LSA, researchers may use more than one weighting scheme, and may subject the results calculated to further factor analysis or clustering. Similarly, in Leximancer, researchers may optionally choose seeded analysis to guide their data exploration, Gaussian instead of linear maps to visualize their results, or may use some of the additional features, such as knowledge pathways or sentiment analysis where appropriate. There is ample literature on each method, in turn, that scholars can refer to. For instance, Stockwell et al. (2009) discuss three case studies on the basis of Leximancer. Kontostathis and Pottenger (2006) discuss the relative performance of the singular value decomposition algorithm settings underlying LSA, and different applications of LSA are discussed in Yeh et al. (2005) or Dumais (2004).

References

ABRAHAMSON E (1996) Management Fashion. *Academy of Management Review* 21(1), 254-285.

AL QENAEI ZM (2009) An Investigation of the Relationship between Consumer Mental Health Recovery Indicators and Clinicians' Reports Using Multivariate Analyses of the Singular Value Decomposition of a Textual Corpus. University of Colorado, Boulder, Colorado.

ANIBA MR, SIGUENZA S, FRIEDRICH A, PLEWNIAK F, POCH O, MARCHLER-BAUER A and THOMPSON JD (2009) Knowledge-based Expert Systems and a Proof-of-Concept Case Study for Multiple Sequence Alignment Construction and Analysis. *Briefings in Bioinformatics* 10(1), 11-23.

BASKERVILLE R and MYERS MD (2002) Information Systems as a Reference Discipline. *MIS Quarterly* 26(1), 1-14.

BASKERVILLE R and MYERS MD (2009) Fashion Waves in Information Systems Research and Practice. *MIS Quarterly* 33(4), 647-662.

BELANGER F, COLLINS RW and CHENEY PH (2001) Technology Requirements and Work Group Communication for Telecommuters. *Information Systems Research* 12(2), 155-176.

BERRISFORD T (1979) Heuristic Development: A Redesign of Systems Design. *MIS Quarterly* 3(1),

BOBROW DG and WHALEN J (2002) Community Knowledge Sharing in Practice: The Eureka Story. *Journal of the Society for Organizational Learning* 4(2), 47-59.

BRADEN B (1989) Developing and Expert Systems Strategy. *MIS Quarterly* 13(4),

BURGESS C and LUND K (1997) Modelling Parsing Constraints with High-dimensional Context Space. *Language and Cognitive Processes* 12(2/3), 177-210.

COUSSEMENT K and VAN DEN POEL D (2008) Improving Customer Complaint Management by Automatic Email Classification Using Linguistic Style Features as Predictors. *Decision Support Systems* 44(4), 870-882.

DAVIES I, GREEN P, ROSEMANN M, INDULSKA M and GALLO S (2006) How do Practitioners Use Conceptual Modeling in Practice? *Data & Knowledge Engineering* 58(3), 358-380.

DE GUINEA AO and MARKUS ML (2009) Why Break the Habit of a Lifetime? Rethinking the Roles of Intention, Habit, and Emotion in Continuing Information Technology Use *MIS Quarterly* 33(3), 433-444.

DONG A (2005) The Latent Semantic Approach to Studying Design Team Communication. *Design Studies* 26(5), 445-461.

DUMAIS S (2004) Latent Semantic Analysis. *Annual Review of Information Science and Technology* 38(1), 188-230.

ELVEVÅG B, FOLTZ PW, WEINBERGER DR and GOLDBERG TE (2007) Quantifying Incoherence in Speech: An Automated Methodology and Novel Application to Schizophrenia. *Schizophrenia Research* 93(1-3), 304-316.

EVANGELOPOULOS N, ZHANG X and PRYBUTOK VR (2011) Latent Semantic Analysis: Five Methodological Recommendations. *European Journal of Information Systems* 20, In Press.

FERNANDEZ WD (2004) Using the Glaserian Approach in Grounded Studies of Emerging Business Practices. *Electronic Journal of Business Research Methods* 2(2), 83-94.

FOLTZ PW (1995) Improving Human-Proceedings Interaction: Indexing the CHI Index. In *CHI 95 Conference Companion: Mosaic of Creativity* (KATZ IR, MACK R, MARKS L, ROSSON MB and NIELSEN J, Eds), pp 101-102, ACM, Denver, Colorado.

GORSUCH RL (1997) Exploratory Factor Analysis: Its Role in Item Analysis. *Journal of Personality Assessment* 68(3), 532-560.

GREMILLION L (1980) Managing the implementation of standardized computer based systems. *MIS Quarterly* 4(4),

HAMPTON JA (1995) Testing the Prototype Theory of Concepts. *Journal of Memory and Language* 34(5), 686-708.

HEPWORTH N and PAXTON SJ (2007) Pathways to Help-Seeking in Bulimia Nervosa and Binge Eating Problems: A Concept Mapping Approach. *International Journal of Eating Disorders* 40(6), 493-504.

HOVORKA DS, LARSEN KRT and MONARCHI DE (2009) Conceptual Convergences: Positioning Information Systems Among the Business Disciplines. In *15th European Conference on Information Systems* (NEWELL S, WHITLEY EA, POULOUDI N, WAREHAM J and MATHIASSEN L, Eds), Uniersity of Verona, Verona, Italy.

INDULSKA M and RECKER J (2008) Design Science in IS Research: A Literature Analysis. In *4th Biennial ANU Workshop on Information Systems Foundations* (GREGOR S and HO S, Eds), ANU E-Press, Canberra, Australia.

JARVENPAA S and LEIDNER D (1998) An Information Company in Mexico: Extending the Resource-Based View of the Firm to a Developing Country Context. *Information Systems Research* 9(4), 342-361.

KING WR (2009) Text Analytics: Boon to Knowledge Management? *Information Systems Management* 26(1), 87.

KONTOSTATHIS A and POTTENGER WM (2006) A Framework for Understanding Latent Semantic Indexing (LSI) Performance *Information Processing & Management* 42(1), 56-73.

KRUSCHKE JK (1992) ALCOVE: An Exemplar-Based Connectionist Model of Category Learning. *Psychological Review* 99(1), 22-44.
LANDAUER TK (2007) LSA As a Theory of Meaning. In *Handbook of Latent Semantic Analysis* (LANDAUER TK, MCNAMARA DS, DENNIS S and KINTSCH W, Eds), pp 3-34, Lawrence Erlbaum Associates, Mahwah, New Jersey.
LANDAUER TK, FOLTZ PW and LAHAM D (1998) Introduction to Latent Semantic Analysis. *Discourse Processes* 25(2 & 3), 259-284.
LARSEN KRT and MONARCHI DE (2004) A Mathematical Approach to Categorization and Labeling of Qualitative Data: The Latent Semantic Categorization Method. *Sociological Methodology* 34(1), 349-392.
LARSEN KRT, MONARCHI DE, HOVORKA DS and BAILEY CN (2008) Analyzing Unstructured Text Data: Using Latent Categorization to Identify Intellectual Communities in Information Systems. *Decision Support Systems* 45(4), 884-896.
LIANG H and XUE Y (2009) Avoidance of Information Technology Threats: A Theoretical Perspective. *MIS Quarterly* 33(1), 71-90.
LOFTIN RD and MOOSBRUKKER J (1982) Organization Development Methods in the Management of the Information Systems Function. *MIS Quarterly* 6(3), 15-28.
MARTIN NJ and RICE JL (2007) Profiling Enterprise Risks in Large Computer Companies Using the Leximancer Software Tool. *Risk Management* 9(3), 188-206.
MCKENNA B and WADDELL N (2007) Media-ted Political Oratory Following Terrorist Events: International Political Responses to the 2005 London Bombing. *Journal of Language and Politics* 6(3), 377-399.
MEADOR C (1984) Setting Priorities for DSS Development. *MIS Quarterly* 8(2),
MILES MB and HUBERMAN M (1994) *Qualitative Data Analysis.* Sage, Thousand Oaks, California.
PENN-EDWARDS S (2010) Computer Aided Phenomenography: The Role of Leximancer Computer Software in Phenomenographic Investigation. *The Qualitative Report* 15(2), 252-267.
PORTER MF (1980) An Algorithm for Suffix Stripping. *Program* 14(3), 130-137.
RAMILLER NC, SWANSON EB and WANG P (2009) Research Directions in Information Systems: Toward an Institutional Ecology. *Journal of the Association for Information Systems* 9(1), 1-22.
RÜLING C-C (2005) Popular Concepts and the Business Management Press. *Scandinavian Journal of Management* 21(2), 177-195.
SALTON G (1989) *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer.* Addison-Wesley, Reading, Massachusetts.
SAUNDERS C (1981) Management Information Systems, Communications, and Departmental Power: An Integrative Model. *Academy of Management Review* 6(3), 431-442.
SIDOROVA A, EVANGELOPOULOS N, VALACICH JS and RAMAKRISHNAN T (2008) Uncovering the Intellectual Core of the Information Systems Discipline. *MIS Quarterly* 32(3), 467-482.
SMITH AE and HUMPHREYS MS (2006) Evaluation of Unsupervised Semantic Mapping of Natural Language with Leximancer Concept Mapping. *Behavior Research Methods, Instruments, and Computers* 38(2), 262-279.
STOCKWELL P, COLOMB RM, SMITH AE and WILES J (2009) Use of an Automatic Content Analysis Tool: A Technique for Seeing Both Local and Global Scope. *International Journal of Human-Computer Studies* 67(5), 424-436.

TAM K (1998) The Impact of Information Technology Investments on Firm Performance and Evaluation: Evidence from Newly Industrialized Economies. *Information Systems Research* 9(1), 85-98.

TRIESCHMANN JS, DENNIS AR, NORTHCRAFT GB and NIEMI AW (2000) Serving Multiple Constituencies in the Business School: MBA Program versus Research Performance. *Academy of Management Journal* 43(6), 1130-1141.

URQUHART C, LEHMANN H and MYERS MD (2010) Putting the Theory Back Into Grounded Theory: Guidelines for Grounded Theory Studies in Information Systems. *Information Systems Journal* 20(4), 357-381.

VESSEY I, RAMESH V and GLASS RL (2002) Research in Information Systems: An Empirical Study of Diversity in the Discipline and Its Journals. *Journal of Management Information Systems* 19(2), 129-174.

WEBER RP (1990) *Basic Content Analysis.* Sage, Newbury Park, California.

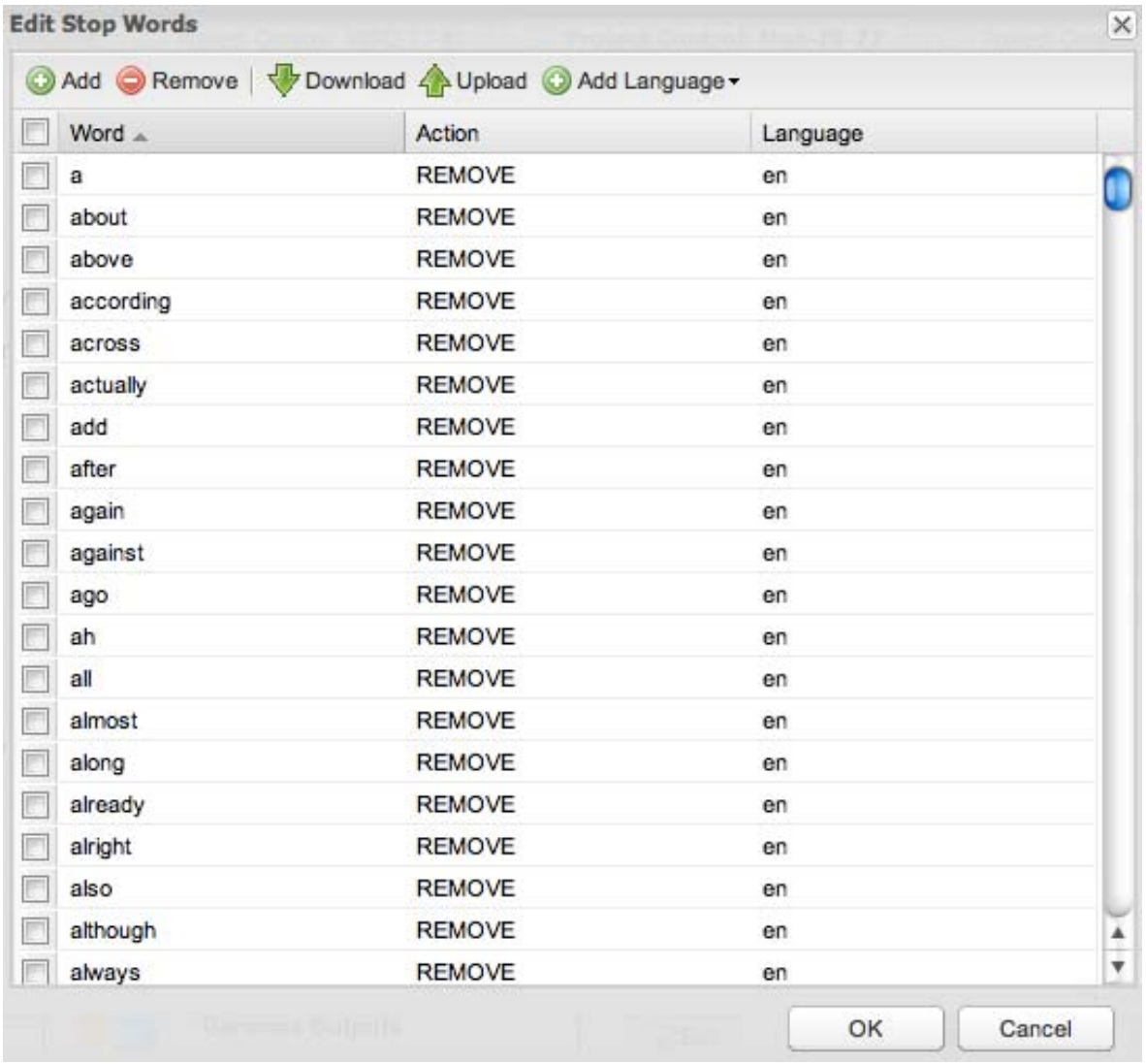
WEI C-P, YANG CC and LIN C-M (2008) A Latent Semantic Indexing-based Approach to Multilingual Document Clustering. *Decision Support Systems* 45(3), 606-620.

YAROWSKY D (1995) Unsupervised Word Sense Disambiguation Rivaling Supervised Methods In *33rd Annual Meeting on Association for Computational Linguistics*, pp 189-196, Association for Computational Linguistics, Cambridge, Massachusetts.

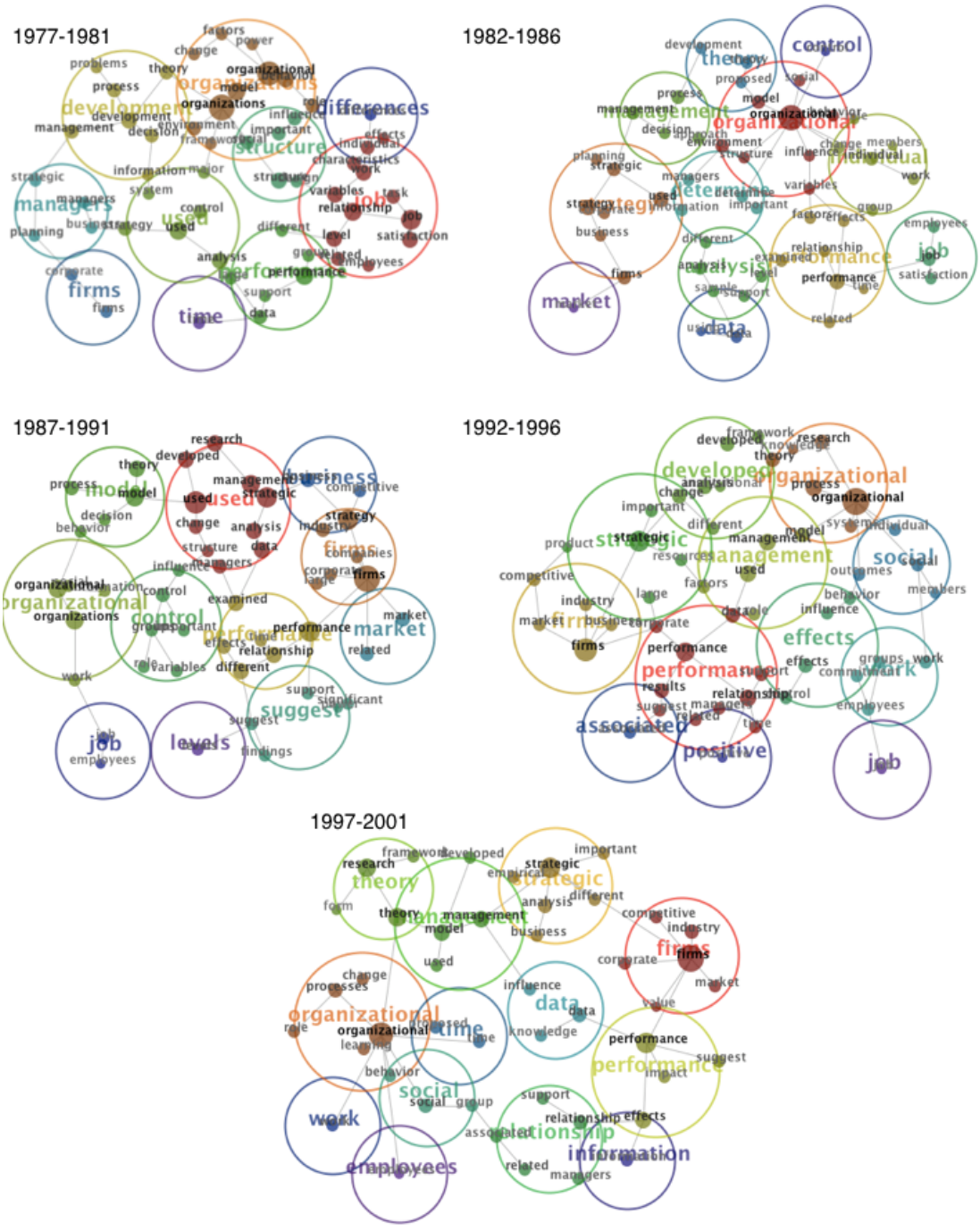
YEH J-Y, KE H-R, YANG W-P and MENG I-H (2005) Text Summarization Using a Trainable Summarizer and Latent Semantic Analysis. *Information Processing & Management* 41(1), 75-95.

ZHAO Y, WANG W and ZHU Y (2010) Antecedents of the Closeness of Human-Avatar Relationships in a Virtual World. *Journal of Database Management* 21(2), 41-68.

Appendix A. Stop word excerpt screenshot from Leximancer 3.5



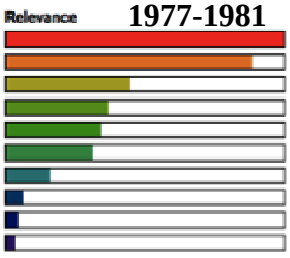
Appendix B – Management Concept Maps 1977-2001



Appendix C – Management Theme Distributions 1977-2001

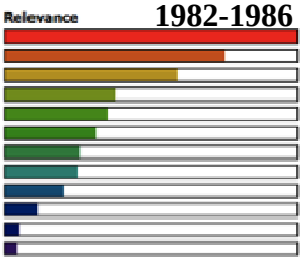
1977-1981

Theme	Connectivity	Relevance
job	100%	
organizations	88%	
development	44%	
used	37%	
performance	34%	
structure	31%	
managers	16%	
firms	06%	
differences	04%	
time	03%	



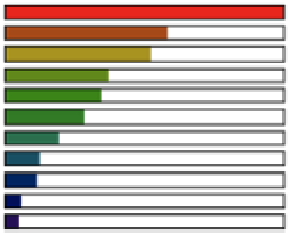
1982-1986

Theme	Connectivity	Relevance
organizational	100%	
strategy	75%	
performance	59%	
individual	38%	
management	35%	
analysis	31%	
job	26%	
determine	25%	
theory	20%	
data	11%	
control	05%	
market	04%	



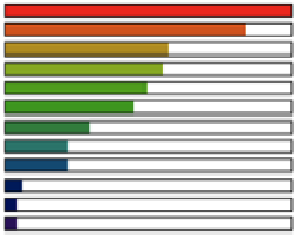
1987-1991

Theme	Connectivity	Relevance
used	100%	
firms	58%	
performance	52%	
organizational	37%	
model	34%	
control	28%	
suggest	19%	
market	12%	
business	11%	
job	05%	
levels	04%	



1992-1996

Theme	Connectivity	Relevance
performance	100%	
organizational	84%	
firms	57%	
management	55%	
developed	49%	
strategic	45%	
effects	29%	
work	22%	
social	22%	
associated	06%	
positive	04%	
job	04%	



1997-2001

Theme	Connectivity	Relevance
firms	100%	
organizational	84%	
strategic	84%	
performance	68%	
theory	65%	
management	62%	
relationship	53%	
social	31%	
data	31%	
time	15%	
work	09%	
information	06%	
employees	06%